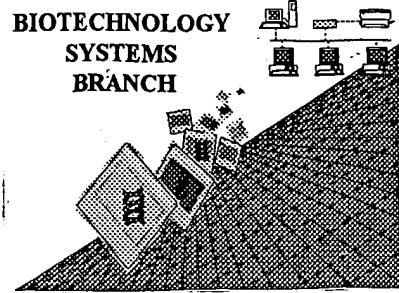


0790

RAW SEQUENCE LISTING ERROR REPORT



The Biotechnology Systems Branch of the Scientific and Technical Information Center (STIC) detected errors when processing the following computer readable form:

Application Serial Number: 10/023,066
Source: OIPE
Date Processed by STIC: 1/15/2002

THE ATTACHED PRINTOUT EXPLAINS DETECTED ERRORS.

PLEASE FORWARD THIS INFORMATION TO THE APPLICANT BY EITHER:

- 1) INCLUDING A COPY OF THIS PRINTOUT IN YOUR NEXT COMMUNICATION TO THE APPLICANT, WITH A NOTICE TO COMPLY or,
- 2) TELEPHONING APPLICANT AND FAXING A COPY OF THIS PRINTOUT, WITH A NOTICE TO COMPLY

FOR CRF SUBMISSION QUESTIONS, PLEASE CONTACT MARK SPENCER, 703-308-4212.

FOR SEQUENCE RULES INTERPRETATION, PLEASE CONTACT ROBERT WAX, 703-308-4216.

PATENTIN 2.1 e-mail help: patin21help@uspto.gov or phone 703-306-4119 (R. Wax)

PATENTIN 3.0 e-mail help: patin3help@uspto.gov or phone 703-306-4119 (R. Wax)

TO REDUCE ERRORED SEQUENCE LISTINGS, PLEASE USE THE **CHECKER VERSION 3.1 PROGRAM**, ACCESSIBLE THROUGH THE U.S. PATENT AND TRADEMARK OFFICE WEBSITE. SEE BELOW FOR ADDRESS:

<http://www.uspto.gov/web/offices/pac/checker>

Applicants submitting genetic sequence information electronically on diskette or CD-Rom should be aware that there is a possibility that the disk/CD-Rom may have been affected by treatment given to all incoming mail.

Please consider using alternate methods of submission for the disk/CD-Rom or replacement disk/CD-Rom, including:

1. EFS-Bio (<<http://www.uspto.gov/ebc/efs/downloads/documents.htm>> , EFS Submission User Manual - ePAVE)
2. U.S. Patent and Trademark Office, Box Sequence, P.O. Box 2327, Arlington, VA 22202
3. Hand Carry directly to:
U.S. Patent and Trademark Office, Technology Center 1600, Reception Area, 7th Floor, Examiner Name,
Sequence Information, Crystal Mall One, 1911 South Clark Street, Arlington, VA 22202
Or
U.S. Patent and Trademark Office, Box Sequence, Customer Window, Lobby, Room 1B03, Crystal Plaza Two,
2011 South Clark Place, Arlington, VA 22202
4. Federal Express, United Parcel Service, or other delivery service to: U.S. Patent and Trademark Office,
Box Sequence, Room 1B03-Mailroom, Crystal Plaza Two, 2011 South Clark Place, Arlington, VA 22202

Raw Sequence Listing Error Summary

ERROR DETECTED	SUGGESTED CORRECTION	SERIAL NUMBER: <u>10/023,066</u>
ATTN: NEW RULES CASES: PLEASE DISREGARD ENGLISH "ALPHA" HEADERS, WHICH WERE INSERTED BY PTO SOFTWARE		
1 <input checked="" type="checkbox"/> Wrapped Nucleic Wrapped Aminos	The number/text at the end of each line "wrapped" down to the next line. This may occur if your file was retrieved in a word processor after creating it. Please adjust your right margin to .3; this will prevent "wrapping."	
2 <input type="checkbox"/> Invalid Line Length	The rules require that a line not exceed 72 characters in length. This includes white spaces.	
3 <input checked="" type="checkbox"/> Misaligned Amino Numbering	The numbering under each 5 th amino acid is misaligned. Do not use tab codes between numbers; use space characters, instead.	
4 <input type="checkbox"/> Non-ASCII	The submitted file was not saved in ASCII(DOS) text, as required by the Sequence Rules. Please ensure your subsequent submission is saved in ASCII text.	
5 <input type="checkbox"/> Variable Length	Sequence(s) _____ contain n's or Xaa's representing more than one residue. Per Sequence Rules, each n or Xaa can only represent a single residue. Please present the maximum number of each residue having variable length and indicate in the <220>-<223> section that some may be missing.	
6 <input type="checkbox"/> PatentIn 2.0 "bug"	A "bug" in PatentIn version 2.0 has caused the <220>-<223> section to be missing from amino acid sequences(s) _____. Normally, PatentIn would automatically generate this section from the previously coded nucleic acid sequence. Please manually copy the relevant <220>-<223> section to the subsequent amino acid sequence. This applies to the mandatory <220>-<223> sections for Artificial or Unknown sequences.	
7 <input type="checkbox"/> Skipped Sequences (OLD RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for each skipped sequence: (2) INFORMATION FOR SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) (i) SEQUENCE CHARACTERISTICS: (Do not insert any subheadings under this heading) (xi) SEQUENCE DESCRIPTION:SEQ ID NO:X: (insert SEQ ID NO where "X" is shown) This sequence is intentionally skipped Please also adjust the "(ii) NUMBER OF SEQUENCES:" response to include the skipped sequences.	
8 <input type="checkbox"/> Skipped Sequences (NEW RULES)	Sequence(s) _____ missing. If intentional, please insert the following lines for each skipped sequence. <210> sequence id number <400> sequence id number 000	
9 <input type="checkbox"/> Use of n's or Xaa's (NEW RULES)	Use of n's and/or Xaa's have been detected in the Sequence Listing. Per 1.823 of Sequence Rules, use of <220>-<223> is MANDATORY if n's or Xaa's are present. In <220> to <223> section, please explain location of n or Xaa, and which residue n or Xaa represents.	
10 <input type="checkbox"/> Invalid <213> Response	Per 1.823 of Sequence Rules, the only valid <213> responses are: Unknown, Artificial Sequence, or scientific name (Genus/species). <220>-<223> section is required when <213> response is Unknown or is Artificial Sequence	
11 <input type="checkbox"/> Use of <220>	Sequence(s) _____ missing the <220> "Feature" and associated numeric identifiers and responses. Use of <220> to <223> is MANDATORY if <213> "Organism" response is "Artificial Sequence" or "Unknown." Please explain source of genetic material in <220> to <223> section. (See "Federal Register," 06/01/1998, Vol. 63, No. 104, pp. 29631-32) (Sec. 1.823 of Sequence Rules)	
12 <input type="checkbox"/> PatentIn 2.0 "bug"	Please do not use "Copy to Disk" function of PatentIn version 2.0. This causes a corrupted file, resulting in missing mandatory numeric identifiers and responses (as indicated on raw sequence listing). Instead, please use "File Manager" or any other manual means to copy file to floppy disk.	
13 <input type="checkbox"/> Misuse of n	n can only be used to represent a single nucleotide in a nucleic acid sequence. N is not used to represent any value not specifically a nucleotide.	

RAW SEQUENCE LISTING
PATENT APPLICATION: US/10/023,066

DATE: 01/15/2002
TIME: 19:01:05

Input Set : A:\BBL037 Sequence Listing.txt
Output Set: N:\CRF3\01152002\J023066.raw

ppr 1,34

SEQUENCE LISTING

- 3 (1) GENERAL INFORMATION:
 - 5 (i) APPLICANT: E. I. DU PONT DE NEMOURS AND
COMPANY
 - 8 (ii) TITLE OF INVENTION: CHIMERIC GENES AND
METHODS FOR INCREASING
delete → INCREASING THE LYSINE
duplicate AND THREONINE CONTENT
OF THE SEEDS OF PLANTS
 - 14 (iii) NUMBER OF SEQUENCES: 107
 - 16 (iv) CORRESPONDENCE ADDRESS:
 - 17 (A) ADDRESSEE: E. I. DU PONT DE NEMOURS
AND COMPANY
 - 19 (B) STREET: 1007 MARKET STREET
 - 20 (C) CITY: WILMINGTON
 - 21 (D) STATE: DELAWARE
 - 22 (E) COUNTRY: U.S.A.
 - 23 (F) ZIP: 19898
 - 25 (v) COMPUTER READABLE FORM:
 - 26 (A) MEDIUM TYPE: FLOPPY DISK
 - 27 (B) COMPUTER: IBM PC COMPATIBLE
 - 28 (C) OPERATING SYSTEM: PC-DOS/MS-DOS
 - 29 (D) SOFTWARE: MICROSOFT WORD VERSION 2.0C
 - 31 (vi) CURRENT APPLICATION DATA:
 - 32 (A) APPLICATION NUMBER: US/10/023,066
 - 33 (B) FILING DATE: 17-Dec-2001
 - 34 (C) CLASSIFICATION:
 - 36 (viii) ATTORNEY/AGENT INFORMATION:
 - 37 (A) NAME: BARBARA C. SIEGELL
 - 38 (B) REGISTRATION NUMBER: 30,684
 - 39 (C) REFERENCE/DOCKET NUMBER: BB-1037-C
 - 41 (ix) TELECOMMUNICATION INFORMATION:
 - 42 (A) TELEPHONE: 302-992-4931
 - 43 (B) TELEFAX: 302-773-0164
 - 44 (C) TELEX: 835420

Does Not Comply
Corrected Diskette Needed
(please try to lengthen
each line - thus, fewer
lines will
be shown)

ERRORED SEQUENCES

- 234 (2) INFORMATION FOR SEQ ID NO: 6:
 - 236 (i) SEQUENCE CHARACTERISTICS:
 - 237 (A) LENGTH: 917 base pairs
 - 238 (B) TYPE: nucleic acid
 - 239 (C) STRANDEDNESS: single
 - 240 (D) TOPOLOGY: linear
 - 242 (ii) MOLECULE TYPE: DNA (genomic)
 - 244 (ix) FEATURE:

P.3

Input Set : A:\BB1037 Sequence Listing.txt
Output Set: N:\CRF3\01152002\J023066.raw

245 (A) NAME/KEY: CDS
246 (B) LOCATION: 3..911
248 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 6:
250 CC ATG GCT ACA GGT TTA ACA GCT AAG ACC GGA GTA GAG CAC TTC GGC 47
251 Met Ala Thr Gly Leu Thr Ala Lys Thr Gly Val Glu His Phe Gly
252 1 5 10 15
254 ACC GTT GGA GTA GCA ATG GTT ACT CCA TTC ACG GAA TCC GGA GAC ATC 95
255 Thr Val Gly Val Ala Met Val Thr Pro Phe Thr Glu Ser Gly Asp Ile
256 20 25 30
258 GAT ATC GCT GCT GGC CGC GAA GTC GCG GCT TAT TTG GTT GAT AAG GGC 143
259 Asp Ile Ala Ala Gly Arg Glu Val Ala Ala Tyr Leu Val Asp Lys Gly
260 35 40 45
262 TTG GAT TCT TTG GTT CTC GCG GGC ACC ACT GGT GAA TCC CCA ACG ACA 191
263 Leu Asp Ser Leu Val Leu Ala Gly Thr Thr Gly Glu Ser Pro Thr Thr
264 50 55 60
266 ACC GCC GCT GAA AAA CTA GAA CTG CTC AAG GCC GTT CGT GAG GAA GTT 239
267 Thr Ala Ala Glu Lys Leu Glu Leu Leu Lys Ala Val Arg Glu Glu Val
268 65 70 75
270 GGG GAT CGG GCG AAG CTC ATC GCC GGT GTC GGA ACC AAC AAC ACG CGG 287
271 Gly Asp Arg Ala Lys Leu Ile Ala Gly Val Gly Thr Asn Asn Thr Arg
272 80 85 90 95
274 ACA TCT GTG GAA CTT GCG GAA GCT GCT GCT TCT GCT GGC GCA GAC GGC 335
275 Thr Ser Val Glu Leu Ala Glu Ala Ala Ser Ala Gly Ala Asp Gly
276 100 105 110
278 CTT TTA GTT GTA ACT CCT TAT TAC TCC AAG CCG AGC CAA GAG GGA TTG 383
279 Leu Leu Val Val Thr Pro Tyr Tyr Ser Lys Pro Ser Gln Glu Gly Leu
280 115 120 125
282 CTG GCG CAC TTC GGT GCA ATT GCT GCA GCA ACA GAG GTT CCA ATT TGT 431
283 Leu Ala His Phe Gly Ala Ile Ala Ala Ala Thr Glu Val Pro Ile Cys
284 130 135 140
286 CTC TAT GAC ATT CCT GGT CGG TCA GGT ATT CCA ATT GAG TCT GAT ACC 479
287 Leu Tyr Asp Ile Pro Gly Arg Ser Gly Ile Pro Ile Glu Ser Asp Thr
288 145 150 155
290 ATG AGA CGC CTG AGT GAA TTA CCT ACG ATT TTG GCG GTC AAG GAC GCC 527
291 Met Arg Arg Leu Ser Glu Leu Pro Thr Ile Leu Ala Val Lys Asp Ala
292 160 165 170 175
294 AAG GGT GAC CTC GTT GCA GCC ACG TCA TTG ATC AAA GAA ACG GGA CTT 575
295 Lys Gly Asp Leu Val Ala Ala Thr Ser Leu Ile Lys Glu Thr Gly Leu
296 180 185 190
298 GCC TGG TAT TCA GGC GAT GAC CCA CTA AAC CTT GTT TGG CTT GCT TTG 623
299 Ala Trp Tyr Ser Gly Asp Asp Pro Leu Asn Leu Val Trp Leu Ala Leu
300 195 200 205
302 GGC GGA TCA GGT TTC ATT TCC GTA ATT GGA CAT GCA GCC CCC ACA GCA 671
303 Gly Gly Ser Gly Phe Ile Ser Val Ile Gly His Ala Ala Pro Thr Ala
304 210 215 220
306 TTA CGT GAG TTG TAC ACA AGC TTC GAG GAA GGC GAC CTC GTC CGT GCG 719
307 Leu Arg Glu Leu Tyr Thr Ser Phe Glu Glu Gly Asp Leu Val Arg Ala
308 225 230 235
310 CGG GAA ATC AAC GCC AAA CTA TCA CCG CTG GTA GCT GCC CAA GGT CGC 767

Input Set : A:\BB1037 Sequence Listing.txt
Output Set: N:\CRF3\01152002\J023066.raw

311 Arg Glu Ile Asn Ala Lys Leu Ser Pro Leu Val Ala Ala Gln Gly Arg
312 240 245 250 255
314 TTG GGT GGA GTC AGC TTG GCA AAA GCT GCT CTG CGT CTG CAG GGC ATC 815
315 Leu Gly Gly Val Ser Leu Ala Ala Lys Ala Leu Arg Leu Gln Gly Ile
316 260 265 270
318 AAC GTA GGA GAT CCT CGA CTT CCA ATT ATG GCT CCA AAT GAG CAG GAA 863
319 Asn Val Gly Asp Pro Arg Leu Pro Ile Met Ala Pro Asn Glu Gln Glu
320 275 280 285
E--> 322 CTT GAG GCT CTC CGA GAA GAC ATG AAA AAA GCT GGA GTT CTA TAA TGAGAATT 918 917
323 Leu Glu Ala Léu Arg Glu Asp Mét Lys Lys Ala Gly Val Léu *
324 290 295 300

1741 (2) INFORMATION FOR SEQ ID NO: 67:

1743 (i) SEQUENCE CHARACTERISTICS:

1744 (A) LENGTH: 28 amino acids
1745 (B) TYPE: amino acid
1746 (C) STRANDEDNESS: unknown
1747 (D) TOPOLOGY: unknown

1749 (ii) MOLECULE TYPE: protein

1751 (ix) FEATURE:
1752 (A) NAME/KEY: Protein
1753 (B) LOCATION: 1..28
1754 (D) OTHER INFORMATION: /label= name

1755 /note= "(SSP 5)4"

1757 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 67:

1759 Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met Glu
1760 1 5 10 15

1762 Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala
E--> 1763 20 20 25 25 misaligned amino acid number

(see item 3 on
Error Summary
Sheet)

1980 (2) INFORMATION FOR SEQ ID NO: 76:

1982 (i) SEQUENCE CHARACTERISTICS:
1983 (A) LENGTH: 175 base pairs
1984 (B) TYPE: nucleic acid
1985 (C) STRANDEDNESS: double
1986 (D) TOPOLOGY: linear

1988 (ii) MOLECULE TYPE: DNA (genomic)

1990 (vi) ORIGINAL SOURCE:
1991 (B) STRAIN: E. coli

1992 (G) CELL TYPE: DH5 alpha

P.4

1994 (vii) IMMEDIATE SOURCE:

1995 (B) CLONE: 5-1

1997 (ix) FEATURE:

1998 (A) NAME/KEY: CDS
1999 (B) LOCATION: 2..172
2000 (D) OTHER INFORMATION: /function= "synthetic

2001 storage protein

2002 /product= "protein"

2003 /gene= "ssp"

2004 /standard_name=

2005 "5.5.5.7.7.7.7.5"

Input Set : A:\BB1037 Sequence Listing.txt
Output Set: N:\CRF3\01152002\J023066.raw

2007 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 76:
2009 C ATG GAG GAG AAG ATG AAG GCG ATG GAG GAG AAG ATG AAG GCG ATG 46
2010 Met Glu Glu Lys Met Lys Ala Met Glu Glu Lys Met Lys Ala Met
2011 1 5 10 15
2013 GAG GAG AAG ATG AAG GCG ATG GAG GAA AAG CTG AAA GCG ATG GAG GAG 94
2014 Glu Glu Lys Met Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu
2015 20 25 30
2017 AAA CTC AAG GCT ATG GAA GAA AAG CTT AAA GCG ATG GAG GAG AAA CTG 142
2018 Lys Leu Lys Ala Met Glu Glu Lys Leu Lys Ala Met Glu Glu Lys Leu
2019 35 40 45
E--> 2021 AAG GCC ATG GAA GAG AAG ATG AAG GCG TGATAG 179/175
2022 Lys Ala Met Glu Glu Lys Met Lys Ala
2023 50 55
2515 (2) INFORMATION FOR SEQ ID NO: 98:
2517 (i) SEQUENCE CHARACTERISTICS:
2518 (A) LENGTH: 59 base pairs
2519 (B) TYPE: nucleic acid
2520 (C) STRANDEDNESS: single
2521 (D) TOPOLOGY: linear
2523 (ii) MOLECULE TYPE: DNA (genomic)
2525 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 98:
E--> 2527 CATGGCGCCC ACCGTGATGA TGGCCTCGTC GGCCACCGCC GTCGCTCCGT TCCAGGGGC 59
W--> 2528 59 → 59
2530 (2) INFORMATION FOR SEQ ID NO: 99:
2532 (i) SEQUENCE CHARACTERISTICS:
2533 (A) LENGTH: 59 base pairs
2534 (B) TYPE: nucleic acid
2535 (C) STRANDEDNESS: single
2536 (D) TOPOLOGY: linear
2538 (ii) MOLECULE TYPE: DNA (genomic)
2540 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 99:
E--> 2542 TTAAGCCCCCT GGAACGGAGC GACGGCGGTG GCCGACGAGG CCATCATCAC GGTGGGCC 59
W--> 2543 59 → 59
2700 (2) INFORMATION FOR SEQ ID NO: 107:
2702 (i) SEQUENCE CHARACTERISTICS:
2703 (A) LENGTH: 29 base pairs
2704 (B) TYPE: nucleic acid
2705 (C) STRANDEDNESS: single
2706 (D) TOPOLOGY: linear
2708 (ii) MOLECULE TYPE: DNA (genomic)
2710 (xi) SEQUENCE DESCRIPTION: SEQ ID NO: 107:
2712 CTCTCGGTAC CTAGTACCTA CTGATCAAC 29
E--> 2714 BB-1037-C → delete

sel item 1
on Env
summary sheet

VERIFICATION SUMMARY
PATENT APPLICATION: US/10/023,066

DATE: 01/15/2002
TIME: 19:01:06

Input Set : A:\BB1037 Sequence Listing.txt
Output Set: N:\CRF3\01152002\J023066.raw

L:32 M:220 C: Keyword misspelled or invalid format, [(A) APPLICATION NUMBER:]
L:33 M:220 C: Keyword misspelled or invalid format, [(B) FILING DATE:]
L:12 M:259 W: Allowed number of lines exceeded, (ii) TITLE OF INVENTION:
L:176 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:1
L:322 M:254 E: No. of Bases conflict, Input:918 Counted:917 SEQ:6
L:776 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:29
L:843 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:31
L:1380 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:53
L:1763 M:332 E: (32) Invalid/Missing Amino Acid Numbering, SEQ ID:67
L:2021 M:254 E: No. of Bases conflict, Input:179 Counted:175 SEQ:76
L:2527 M:254 E: No. of Bases conflict, Input:0 Counted:59 SEQ:98
L:2528 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:98
L:2542 M:254 E: No. of Bases conflict, Input:0 Counted:59 SEQ:99
L:2543 M:336 W: Invalid Amino Acid Number in Coding Region, SEQ ID:99
L:2638 M:341 W: (46) "n" or "Xaa" used, for SEQ ID#:104
L:2714 M:334 W: (2) Invalid Amino Acid in Coding Région, NUMBER OF INVALID KEYS:1
L:2714 M:333 E: Wrong sequence grouping, Amino acids not in groups!